

VU Research Portal

Kwaliteitscriteria voor ontwerpgericht wetenschappelijk onderzoek

van Burg, J.C.

published in

Handboek ontwerpgericht wetenschappelijk onderzoek: Wetenschap met effect
2011

[Link to publication in VU Research Portal](#)

citation for published version (APA)

van Burg, J. C. (2011). Kwaliteitscriteria voor ontwerpgericht wetenschappelijk onderzoek. In J. van Aken, & D. Andriessen (Eds.), *Handboek ontwerpgericht wetenschappelijk onderzoek: Wetenschap met effect* (pp. 146-164). Boom Lemma.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Hoofdstuk 9: Kwaliteitscriteria voor ontwerpgericht wetenschappelijk onderzoek¹

Elco van Burg

Vrije Universiteit, Amsterdam

‘The proof of the pudding is the eating’. De kwaliteit van het bakproces moet je beoordelen door het resultaat te proeven. Zo is het ook met een ontwerp. De beste test voor een ontwerp is de invoering ervan. Evenals bij het bakken van een taart wil je echter liever toch al iets eerder weten of het wat op gaat leveren. Een goed recept, de goede ingrediënten, de juiste temperatuur maken het waarschijnlijk dat je een goed resultaat krijgt.

Voor ontwerpgericht onderzoek is het daarom van belang om de kwaliteitscriteria voor, tijdens en na het onderzoek toe te passen (Morse, Barrett, Mayan, Olson, & Spiers, 2008). Dus niet alleen na afloop van het bakken de taart proeven, maar vooraf ook kijken of het recept wel goed is en of alle andere factoren ook in orde zijn. In dit hoofdstuk wil ik eerst een overzicht geven van kwaliteitscriteria voor onderzoek in het algemeen. Daarna bespreek ik de specifieke invulling van deze kwaliteitscriteria voor ontwerpgericht onderzoek en introduceer ik ook specifieke maatregelen om de kwaliteit van het onderzoek te waarborgen. In het laatste deel van dit hoofdstuk geef ik een illustratie van het functioneren van deze ontwerpcriteria in een onderzoek naar het stimuleren van spin-offs van universiteiten.

Algemene kwaliteitscriteria

Een kwalitatief hoogstaand onderzoek is *valide* (ofwel geldig) en *betrouwbaar*. Deze criteria gelden voor alle soorten onderzoek, dus ook voor ontwerpgericht gericht onderzoek. Wel hangt de specifieke invulling van deze criteria af van het type onderzoek. Validiteit en betrouwbaarheid worden in kwantitatief onderzoek op een andere manier gewaarborgd dan in kwalitatief onderzoek.

¹ Gepubliceerd als: Van Burg, E. (2011). Kwaliteitscriteria voor ontwerpgericht wetenschappelijk onderzoek. In J. van Aken & D. Andriessen (Eds.), *Handboek ontwerpgericht wetenschappelijk onderzoek: Wetenschap met effect* (pp. 146-164). Den Haag: Boom Lemma.

Betrouwbaarheid

Betrouwbaarheid betreft de consistentie en stabiliteit van de uitkomsten van het onderzoek (Lee, 1999). Krijg ik dezelfde resultaten als ik het onderzoek nog een keer uitvoer? En krijgt iemand anders ook deze resultaten als het onderzoek nog een keer uitgevoerd wordt? Worden dezelfde dingen gemeten als het onderzoek met andere methoden uitgevoerd wordt? Je kunt het vergelijken met een auto. Als de auto altijd start is hij betrouwbaar, maar als hij onverwacht af en toe het niet doet, is hij onbetrouwbaar.

In kwantitatief vragenlijstonderzoek gaat het om de consistentie van de vragen onderling (laten alle vragen hetzelfde patroon zien?) en de stabiliteit van de resultaten bij het herhalen van het onderzoek (test-hertest stabiliteit) (Lee, 1999). De mate van overeenkomst tussen vergelijkbare antwoorden en tussen een eerste en een herhaalde vragenlijst geeft de betrouwbaarheid aan. In kwantitatief onderzoek wordt de interne consistentie van een vragenlijst uitgedrukt met een coëfficiënt zoals Cronbach's Alpha . In kwalitatief onderzoek werkt dit op vergelijkbare manier (Lee, 1999). Een interview waarin iemand steeds andere antwoorden geeft op vergelijkbare vragen is onbetrouwbaar. En bij herhaling van een interview met dezelfde persoon verwacht je dat iemand ongeveer dezelfde antwoorden geeft of bij afwijkende antwoorden aan kan geven waarom hij of zij er nu anders over denkt.

Validiteit

Bij validiteit gaat het om de vraag of de gehanteerde procedures het mogelijk maken om geldige uitspraken te doen over het object van het onderzoek. Met andere woorden: geven mijn uitspraken werkelijk weer wat er aan de hand is? Validiteit wordt meestal onderverdeeld in verschillende soorten.

Construct validiteit betreft de mate waarin een meetinstrument het bedoelde fenomeen weergeeft (Lee, 1999). Meet je werkelijk wat je bedoelde te meten of meet je iets anders? Een construct is de representatie die de onderzoeker maakt van een bepaald fenomeen. De construct validiteit van het meetinstrument geeft aan in hoeverre het fenomeen en het construct met elkaar overeenkomen. Er zitten twee kanten aan construct validiteit (Van Aken, Berends, & van der Bij, 2007). Ten eerste moet het construct goed gedefinieerd zijn en ten tweede moet het meetinstrument dit construct in

voldoende mate overlappen en geen andere dingen meten. Als je bijvoorbeeld de reputatie van een manager wilt meten en je vraagt aan een willekeurige selectie van werknemers of deze manager in staat is om de organisatie te besturen, meet je niet alleen reputatie, maar ook aspecten van vertrouwen in de manager en een inschatting van de capaciteiten van de manager. Dit onderzoek heeft dan een lage construct validiteit, die verbeterd kan worden door eerst het begrip reputatie goed te definiëren en een daarop aansluitend meetinstrument te ontwikkelen.

Convergente validiteit geeft aan dat verschillende metingen of waarnemingen van hetzelfde construct dezelfde resultaten zouden moeten geven (Lee, 1999). Wanneer interviews met meerdere mensen op dezelfde interviewvragen overeenkomstige dingen laten zien, is er convergentie (Yin, 2003). Aan de andere kant vereist *discriminerende validiteit* dat metingen of waarnemingen van verschillende constructen ook verschillende resultaten moeten opleveren (Lee, 1999). Als je de reputatie van de manager wilt meten en ook het vertrouwen in de manager, dan geeft het een bewijs van discriminerende validiteit als je dan ook verschillende antwoorden vindt.

Interne validiteit of *criterium validiteit* betreft de vraag of er werkelijk een relatie is tussen twee of meer constructen die gemeten zijn (Lee, 1999; Van Aken et al., 2007). Bijvoorbeeld: leidt A tot B of leidt B tot A? Of is er nog een factor C in het spel die zowel A als B verklaart? Als we waarnemen dat jongeren vaker fietsen dan ouderen, zouden we kunnen zeggen dat leeftijd ons fietsgedrag bepaalt. Het is echter zeer goed mogelijk dat een derde factor, namelijk het in bezit zijn van een auto en een rijbewijs, de verklaring vormt.

Externe validiteit geeft aan in hoeverre de onderzoeksresultaten van een onderzoek in een specifieke context of in een specifieke groep *gegeneraliseerd* kunnen worden naar andere contexten of naar een grotere groep (Lee, 1999). *Analytische generaliseerbaarheid* betreft het generaliseren van de onderzoeksresultaten op basis van een aantal specifieke kenmerken. Als het stimuleren van ondernemerschap op universiteiten onderzocht wordt, kunnen de resultaten gegeneraliseerd worden naar vergelijkbare universiteiten. De analytische generaliseerbaarheid van de resultaten kan dan beoordeeld worden door een vergelijking van de essentiële kenmerken van de onderzochte universiteiten met die van de groep universiteiten waarvoor de geldigheid van die resultaten geclaimd wordt. Die kenmerken zijn dan bijvoorbeeld omvang en onderzoeksfocus en een overeenkomstige wetgeving. Uit het bepalen van de analytische generaliseerbaarheid volgt de mate van transfereerbaarheid van de resultaten (Lee, 1999; Lincoln & Guba, 1985). Als essentiële kenmerken uit de onderzochte context ook

in een andere situatie aanwezig zijn, mag aangenomen worden dat de resultaten ook transfereerbaar zijn naar die situatie (zie ook hoofdstuk 2 over de transfereerbaarheid van onderzoeksresultaten).

Het onderschrijven van deze criteria is overigens afhankelijk van de visie op de werkelijkheid. In de ontologische positie van een 'realist' zijn deze criteria geldig, omdat het object van het onderzoek bestaat, los van de observator. Als deze aanname losgelaten wordt, hebben verschillen in waarnemingen weinig waarde meer en is het niet mogelijk om een representatie van de 'werkelijkheid' te maken. Het waargenomen object hangt dan onlosmakelijk samen met de waarnemer en dit object is zelf ook zelf veranderlijk (Guba & Lincoln, 1994).

Kwaliteitscriteria in ontwerpgericht onderzoek

Voor ontwerpgericht onderzoek krijgen de algemene criteria van betrouwbaarheid en validiteit specifieke invullingen. Belangrijk is dat deze criteria zowel over het onderzoeksproces als over de uitkomst van dat onderzoeksproces gaan (Morse et al., 2008). In de opzet van een ontwerpgericht onderzoek moet je maatregelen nemen om validiteit en betrouwbaarheid te bevorderen. Nadat het onderzoek klaar is, dienen de criteria om vast te stellen wat de kwalitatieve zeggingskracht is van het onderzoek. Tabel 1 geeft een overzicht van de kwaliteitscriteria met richtlijnen voor de toepassing ervan in ontwerpgericht onderzoek.

Betrouwbaarheid

Voor de betrouwbaarheid van ontwerpgericht onderzoek is vooral stabiliteit en navolgbaarheid van belang. Een strikte toets voor stabiliteit is het herhalen van het onderzoek. Letterlijke replicatie is echter meestal niet mogelijk, al is het alleen maar omdat het niet meer de eerste keer is dat betrokkenen aan dit onderzoek meewerken (Wester & Peters, 2004). Tussen verschillende bronnen en gevallen kan echter wel replicatie van het onderzoek plaatsvinden. Als de resultaten laten zien dat de uitkomsten bij dezelfde methode en in vergelijkbare gevallen stabiel zijn, geeft dat aan dat de resultaten betrouwbaar zijn.

Meer op een detailniveau kan stabiliteit van de bevindingen vastgesteld worden door meerdere keren een gegevensbron te onderzoeken, bijvoorbeeld door het herhalen van een interview. De resultaten zullen niet 100% hetzelfde zijn, maar onverklaarbare

inhoudelijke verschillen mogen er ook niet optreden. In het verlengde hiervan kunnen 'member checks' gedaan worden om de stabiliteit van de resultaten vast te stellen (Swanborn, 1996). Bij een member check, zie ook hoofdstuk 15, worden de mensen die geïnterviewd zijn of die de gegevens aangeleverd gevraagd of ze de onderzoeksresultaten willen controleren. Klopt het? Kunnen ze zich erin vinden? Als dat het geval is, geeft het aan dat de interpretatie correct is verlopen, maar ook geeft het een indicatie van de stabiliteit van de onderzoeksresultaten.

Omdat strikte replicatie van een onderzoek door een andere onderzoeker meestal niet mogelijk is, is het criterium voor een betrouwbaar onderzoek vooral dat het navolgbaar is. Het zou op een vergelijkbare manier uitgevoerd moeten kunnen worden door een andere onderzoeker. Een uitgewerkt en gedetailleerd onderzoeksprotocol is essentieel voor de navolgbaarheid (Yin, 2003). Dit protocol moet zoveel als mogelijk is van tevoren ontwikkeld zijn en ook alle extra stappen die gedurende het onderzoek genomen worden, moeten erin opgenomen worden. Het is aan te raden om dit protocol als werkdocument regelmatig aan te vullen. Belangrijke onderdelen van dit protocol zijn (Yin, 2003): de doelen en onderwerpen van het onderzoek, de afbakening van de specifieke onderzoekscontext, de gegevensbronnen, de betrokken onderzoekers en hun rollen, de vragen voor de dataverzameling, de analysemethoden en de rapportagemethoden. In de opzet van dit protocol is het van belang om de verschillende aspecten van betrouwbaarheid en validiteit mee te nemen.

Veel ontwerpgericht onderzoek gebruikt naast eigen veldonderzoek ook bestaand onderzoek. Door een grondige en systematische synthese kan zo ontwerp-kennis uit bestaand onderzoek gedestilleerd worden (Denyer & Tranfield, 2006; Denyer, Tranfield, & Van Aken, 2008), zie ook hoofdstuk 11. Dit geeft *geadopteerde betrouwbaarheid*: de betrouwbaarheid van de in de literatuur beschreven onderzoeken bepaalt de betrouwbaarheid van de uitkomsten van de synthese. Daarom is een goede selectie van literatuur essentieel. Bij een synthese kan een spreekwoordelijke rotte appel de betrouwbaarheid van de ontwerp-kennis behoorlijk verlagen. Om de betrouwbaarheid van zulke op bestaande literatuur gebaseerde ontwerp-kennis te waarborgen is een goed onderzoeksprotocol erg belangrijk (Tranfield, Denyer, & Smart, 2003). Hier gelden opnieuw de criteria van stabiliteit en herhaalbaarheid. Als iemand anders literatuur gaat verzamelen met dit protocol, en daar dezelfde uitkomsten uit komen, is deze ontwerp-kennis betrouwbaar (binnen de grenzen van de betrouwbaarheid van de onderzoeken die in de synthese zijn opgenomen).

Validiteit

In kwalitatief onderzoek, wat ontwerpgericht onderzoek meestal is, kan *construct validiteit* gewaarborgd worden door te zorgen voor een duidelijke bewijsketen (Lee, 1999; Yin, 2003). Dit uit de criminologie afkomstige begrip drukt uit dat alle onderdelen in de bewijsketen op elkaar moeten aansluiten. Dat betekent dat je de verschillende onderdelen in het onderzoek van een construct met elkaar samen moeten hangen en logisch op elkaar moeten aansluiten. Als bijvoorbeeld verschillend bronnen gebruikt worden voor het meten van een construct, moet aangegeven worden hoe die informatie leidt tot een bepaalde interpretatie van het construct en hoe daar de resultaten uit volgen. Overigens kunnen de effecten van interventies in ontwerpgericht onderzoek vaak ook kwantitatief gemeten worden. Daarbij moet je uiteraard nagaan of de construct validiteit van deze effecten voldoende is. Dat kan door bijvoorbeeld meetinstrumenten te gebruiken waarvan de construct validiteit al is vastgesteld.

De beste strategie voor het bereiken van *convergente en discriminerende validiteit* bestaat uit het gebruiken van meerdere, ongelijksoortige gegevensbronnen in het ontwerpgerichte onderzoek en onderzoeken of dezelfde constructen eendere effecten hebben en verschillende constructen verschillende effecten (Lee, 1999; Yin, 2003). Een dergelijke verzameling van gegevens kan bijvoorbeeld bestaan uit interviews, e-mails, participant-observaties, krantenartikelen en financiële gegevens. Dit is het principe van triangulatie: het combineren van meerdere bronnen en methoden om verschillende facetten van het onderwerp van onderzoek te belichten en op elkaar te betrekken (Jick, 1979). Inhoudelijke verschillen tussen bronnen kunnen systematisch geanalyseerd worden door codeertechnieken te gebruiken (Miles & Huberman, 1994; Wester & Peters, 2004). Zo worden uitspraken of stukken van een document, figuur, plaatje of video gelabeld met een 'code'. Door codes met elkaar te vergelijken en tegen elkaar uit te zetten ontstaat inzicht in de patronen die in de gegevens aanwezig zijn (Strauss & Corbin, 1998). Een goed onderzoek bereikt dan op een gegeven moment convergentie in de onderzoekresultaten. Vergelijkbare bronnen (met kenmerken die gaandeweg duidelijk worden) laten hetzelfde patroon zien. Aan de andere kant is er dan sprake van divergentie: verschillende bronnen geven verschillende patronen. Dan is er verzadiging bereikt en kan in principe het verzamelen van extra materiaal gestopt worden (Van Aken et al., 2007).

Voor *interne validiteit* is de interne samenhang in de onderzoeksresultaten belangrijk. Het doel van ontwerpgericht onderzoek is om ontwerpkenis, bijvoorbeeld in de vorm

van ontwerpstellingen of ontwerpprincipes op te leveren. Een goed ontwerpprincipe geeft een korte verklaring van waarom een bepaalde interventie in een specifieke context tot een waarschijnlijke uitkomst leidt. Als de interne logica van een ontwerpprincipe overtuigend, helder en consistent is, is er sprake van een vorm van interne validiteit. Deze interne validiteit kan ook getoetst worden, wat een sterkere bevestiging (of verwerping) van de interne validiteit geeft. Yin (2003) geeft hiervoor drie technieken. Het onderzoeken van alternatieve verklaringen (*rival explanations*) voor de beschreven relatie geeft aan of de voorgestelde verklaring de beste verklaring is of niet, zie ook hoofdstuk 16. Als tweede strategie kan het onderzoek naar het voorkomen van het patroon (van bijvoorbeeld de CIMO stelling) in andere bronnen laten zien dat hetzelfde patroon aanwezig is (*pattern matching*). Dat geeft een indicatie dat de interne validiteit goed is. Een nog sterkere vorm hiervan is, ten derde, dat het patroon dat in één setting of case geobserveerd is ook in andere cases voorkomt (*case replication*).

De *externe validiteit* van een ontwerpgericht onderzoek is belangrijk om te kunnen bepalen of de ontwerp kennis niet alleen in de onderzochte context geldig is, maar ook in andere contexten. Met andere woorden: zijn de onderzoeksresultaten *transfereerbaar* naar andere contexten? (zie hoofdstuk 2). Tijdens de onderzoeksopzet dient nagedacht te worden over de beoogde externe validiteit. Eerst moet dan vastgesteld worden welke kenmerken van belang zijn voor het onderzoek en het ontwerp. Dat kunnen voor een bedrijfskundig ontwerp bijvoorbeeld de wet- en regelgeving van de overheid zijn, de grootte van het bedrijf, de industrie, de leeftijd en de soort producten of diensten die het bedrijf heeft. Door gevallen te selecteren die de gewenste kenmerken hebben, wordt de externe validiteit bepaald. Als het onderzoek goed uitgevoerd wordt, zijn de resultaten ook toepasbaar voor andere bedrijven die de kenmerken hebben van de gekozen gevallen. Of dat ook echt het geval is, kan getoetst worden door, nadat de ontwerp kennis is gecodificeerd in ontwerpprincipes of zelfs in een uitgewerkt ontwerp, het ontwerp toe te passen in een onafhankelijk geval met de juiste kenmerken. In een wat zwakkere vorm kan een review van het ontwerp door experts die niet bij het onderzoek betrokken waren ook een validatie geven van de geldigheid van het ontwerp buiten de context waarin het onderzoek is gedaan.

Pragmatische validiteit

Specifiek voor ontwerpgericht onderzoek is pragmatische validiteit van belang (Romme, 2003; Van Aken & Romme, 2009). Pragmatische validiteit betreft de mate waarin het

onderzoek richtlijnen oplevert die bij het daadwerkelijk toepassen van die richtlijnen (binnen het aangegeven toepassingsdomein) de gewenste uitkomsten geven (Kvale & Brinkmann, 2009). Zijn de uitkomsten van het onderzoek richtinggevend voor handelingen die het gewenste resultaat opleveren? Het onderzoek zelf ging over een bestaande situatie of een situatie in het verleden; pragmatische validiteit betreft de mate waarin de onderzoeksresultaten leiden tot handelingen die in de toekomst de gewenste effecten opleveren. Uiteraard moeten deze onderzoeksresultaten zelf ook van hoge kwaliteit zijn; een slecht en onbetrouwbaar onderzoek kan alleen 'per ongeluk' tot goede en effectieve interventies leiden. Onderzoek met een hoge mate van pragmatische validiteit levert kennis op die helpt om praktijken te verbeteren (Taylor, 1985). Pragmatische validiteit heeft ook een belangrijke ethische kant (Miles & Huberman, 1994). Wie en wat wordt beter van het onderzoek? Of zijn er partijen die door de resultaten benadeeld worden? Het is belangrijk om na te gaan dat onderzoeksresultaten niet alleen gebruikt worden, maar ook *misbruikt* kunnen worden (zie paragraaf 6.4).

Pragmatische validiteit vereist dat in het onderzoek gedrag en handelingen en de uitkomsten daarvan onderzocht worden (Kvale & Brinkmann, 2009). Als het onderzoek alleen uit interviews of documentenanalyse bestaat waarin de actoren hun eigen handelen en de effecten daarvan beschrijven, kan de pragmatische validiteit binnen het onderzoek laag zijn. Er kan namelijk een groot verschil zijn tussen de 'theorie' die iemand onderschrijft en het daadwerkelijke handelen van die persoon (Argyris, 2006). Daarom is het beter om interviews te combineren met gedragsobservaties of observaties als participant (Van Burg, Romme, Gilsing, & Reymen, 2008). Als een wat zwakkere vorm van pragmatische validiteit kunnen hier ook 'think aloud' protocollen gebruikt worden (Ericsson & Simon, 1993; Plsek, Bibby, & Whitby, 2007), waarbij scenario's aan veldexperts voorgelegd worden en ze hardop moeten bedenken welke beslissing ze zouden nemen om tot een gewenst resultaat te komen.

Een met pragmatische validiteit sterk samenhangend kwaliteitscriterium is *bruikbaarheid*: is de kennis die het ontwerpgerichte onderzoek oplevert bruikbaar voor de beoogde doelgroep om gedrag en handelingen te veranderen (Dijk, De Goede, 't Hart, & Teunissen, 1991; zie ook hoofdstuk 13). Helpt de kennis om gerichte interventies uit te voeren? Hoe praktischer de resultaten van een ontwerpgericht onderzoek zijn, hoe meer dit de pragmatische validiteit van die resultaten ondersteunt. Ontwerpprincipes kunnen nog redelijk algemeen zijn. Het ontwerp dat op basis van deze ontwerpregels gemaakt wordt, maakt pas echt duidelijk of het onderzoek voldoende pragmatische validiteit bezit. Is dit ontwerp daadwerkelijk toepasbaar en zo ja, geeft de beschrijving van het ontwerp voldoende handvatten voor deze toepassing?

Tabel 9.1 Kwaliteitscriteria voor ontwerpgericht onderzoek

Criterium	Definitie	Uitvoering in ontwerpgericht onderzoek
<i>Betrouwbaarheid</i>	<i>De mate waarin het onderzoek stabiel en consistent is over de tijd en tussen verschillende onderzoekers en methoden.</i>	
Consistentie	De mate waarin de onderzoeksresultaten onderling vergelijkbare uitkomsten geven.	Replicatie bij verschillende cases
Stabiliteit	De mate waarin de onderzoeksresultaten stabiel blijven.	<ul style="list-style-type: none"> - Herhalen van onderzoek van een gegevensbron - Member checks - Navolgbaarheid van onderzoeksprotocol Bij literatuursynthese: <ul style="list-style-type: none"> - Betrouwbaarheid onderliggende studies - Navolgbaarheid review protocol
<i>Validiteit</i>	<i>De mate waarin gehanteerde procedures het mogelijk maken om geldige uitspraken te doen over het onderzoeksobject.</i>	
Construct validiteit	De mate waarin een meetinstrument het bedoelde 'construct' meet.	Zorgen voor duidelijke bewijsketen in de meting van het construct.
Convergente validiteit	De mate waarin metingen van hetzelfde construct dezelfde resultaten geven.	<ul style="list-style-type: none"> - Meerdere databronnen, triangulatie - Systematisch coderen
Discriminerende validiteit	De mate waarin metingen van verschillende constructen verschillende resultaten geven.	<ul style="list-style-type: none"> - Meerdere databronnen, triangulatie - Systematisch coderen
Interne validiteit	De mate waarin er een werkelijke relatie is tussen twee constructen die gemeten zijn.	<ul style="list-style-type: none"> - Logica en consistentie van verklaring (in CIMO logica) Onderzoeken van: <ul style="list-style-type: none"> - rivaliserende verklaringen - herhaling van het patroon - herhaling in andere cases
Externe validiteit	De mate waarin resultaten toepasbaar zijn in een grotere groep of in andere contexten.	<ul style="list-style-type: none"> - Bepalen van kenmerken op basis waarvan gegeneraliseerd kan worden. - Toepassen in een ander geval. - Expert review.
Pragmatische validiteit	De mate waarin het onderzoek richtlijnen oplevert die bij het daadwerkelijk toepassen van die richtlijnen (binnen het aangegeven toepassingsdomein) de gewenste uitkomsten geven.	<ul style="list-style-type: none"> - Onderzoeken van handelingen of gebruik maken van 'think aloud' protocollen. - Heldere ontwerpprincipes - Gebruik maken van verhalen - Gebruik maken van visualisaties - Testen van het ontwerp

De bruikbaarheid voor de voor de beoogde doelgroep kan verhoogd worden door de presentatie van de uitkomsten van het onderzoek goed op deze doelgroep af te te

stemmen. De pragmatische validiteit van ontwerpstellingen kan vastgesteld worden door ze te testen (leidt A, door middel van B, tot C?), door de begrippen goed en helder te operationaliseren en door te beschrijven hoe vanuit deze ontwerpprincipes een ontwerp ontwikkeld kan worden (Worren, Moore, & Elliott, 2002). De pragmatische validiteit van een ontwerp, op basis van de ontwerpprincipes, kan verhoogd worden door praktische, plausibele en overtuigende verhalen aan te bieden die beschrijven hoe het in de praktijk werkt of werkte (Worren et al., 2002). Visuele representaties van het ontwerp kunnen helpen om de bruikbaarheid te vergroten, omdat een plaatje vaak meer zegt dan duizend woorden. Een heldere figuur kan precies de logica samenvatten die de gebruiker nodig heeft om het ontwerp in zijn eigen specifieke context te implementeren (Romme & Endenburg, 2006; Simon, 1996).

Testen van ontwerpen

Werkt het? Dat is de ultieme test voor de pragmatische validiteit van ontwerpgericht onderzoek. Leidt een generiek ontwerp, gebaseerd op de gepresenteerde ontwerpprincipes en ingevoerd in een praktijksituatie echt tot de gewenste verbetering? Het testen van een ontwerp kan op verschillende manieren, variërend van testen die weinig tijd en capaciteit kosten, tot het herhaaldelijk implementeren van het hele ontwerp (zie ook hoofdstuk 10 over het valideren van ontwerp kennis). Uiteraard levert deze laatste praktijktest de sterkste bevestiging of verwerping van de pragmatische validiteit van het ontwerp op. Maar dat brengt wel kosten met zich mee. Als in de implementatie van een ontwerp blijkt dat het ontwerp toch niet werkt, is niet alleen het ontwerp niet bruikbaar, maar zijn meestal ook tijd, geld en moeite verloren gegaan met het invoeren van het ontwerp. Daarom is het goed om verschillende testen van het ontwerp te overwegen.

Een *parameter-specificatie* kan dienen om een model van het ontwerp tot in de details uit te werken (Suh, 2001). Aan de hand van alle parameters die in het ontwerp een rol hebben, kan dan gekeken worden of er logische tegenstrijdigheden zijn en of het ontwerp inderdaad zo zal functioneren als het bedoeld is. Met een *design-structure matrix* kan vervolgens nagegaan worden wat de interne afhankelijkheden zijn binnen het ontwerp (Baldwin & Clark, 2000). Deze parameter-gebaseerde benadering dwingt om alle relaties tussen delen van het ontwerp te bekijken. Dit laat bijvoorbeeld zien dat het logisch onmogelijk is om in het ontwerp voor te schrijven dat een innovatie-organisatie uit zelfsturende teams moet bestaan en tegelijkertijd aan te bevelen dat alle teams volgens tevoren vastgelegde protocollen werken.

Het beschrijven van *scenario's* en het analyseren van het prestaties van het ontwerp in ieder van die scenario's kan duidelijk maken wanneer het ontwerp wel en wanneer het niet goed functioneert. Voor het ontwikkelen van scenario's moeten eerst een aantal waarschijnlijke ontwikkelingen, trends of gebeurtenissen geïdentificeerd worden. Bijvoorbeeld in een bedrijfscontext: het maken van verlies, het falen van een belangrijk product of een reorganisatie van het bedrijf. Door verschillende combinaties te maken van het aanwezig en afwezig zijn van deze ontwikkelingen, ontstaan dan een aantal scenario's (Dammers, 2000). Vervolgens moet dan systematisch nagegaan worden – het liefst met verschillende betrokkenen of veldexperts – hoe het ontwerp waarschijnlijk zal presteren in ieder van deze scenario's.

Een ontwerp kan ook getest worden door het te *modelleren* in een formeel model, bijvoorbeeld in een System Dynamics (SD) model of een kwantitatief variantiemodel, en vervolgens het gedrag of de uitkomsten van het model te bestuderen. De kracht van dergelijke formele modellen is dat de onderlinge afhankelijkheden in het model gespecificeerd moeten worden, met als gevolg dat eventuele lacunes of tegenstrijdigheden duidelijk worden. Een SD model maakt het mogelijk om het dynamische gedrag te onderzoeken (Stermann, 2000). Zo kunnen bijvoorbeeld scenario's opgenomen worden en kan gekeken worden wat de verschillende effecten zijn die het ontwerp in de praktijk zou kunnen teweegbrengen.

Een *gebruikers-review* van het ontwerp levert ten eerste een validatie op of het ontwerp begrepen wordt door de gebruikers. Ook kunnen deze gebruikers reflecteren op de vraag of en hoe het ontwerp zou werken. Wat levert het op? Wat zijn de nadelen die ze zien?

Een *pilot test* is een kleinschalige invoering van het ontwerp. Als het bijvoorbeeld een ontwerp voor teamsamenwerking in een hele ziekenhuisorganisatie betreft, kan de pilottest bijvoorbeeld in één afdeling plaatsvinden. Zo blijven de kosten en risico's beperkt, terwijl wel duidelijk wordt of het ontwerp de gewenste resultaten oplevert.

De volgende stap is dan het daadwerkelijk *invoeren* van het ontwerp, op de gewenste schaal, door de ontwerper of mensen die bij het ontwerp betrokken waren. Hier kunnen de resultaten van de 'test' echter nog beïnvloed worden door de ontwerper(s) en onderzoeker(s). Daarom is de beste validatie van de pragmatische validiteit van het ontwerp het uitvoeren van een *beta-test* (zie hoofdstuk 14). Hierbij wordt het ontwerp ingevoerd door iemand die niet bij het onderzoek en het maken van het ontwerp betrokken was (Stam, 2007). Een positieve uitkomst van deze test laat zien dat het ontwerp zelf de gewenste verbetering teweeg brengt, los van de betrokken onderzoekers.

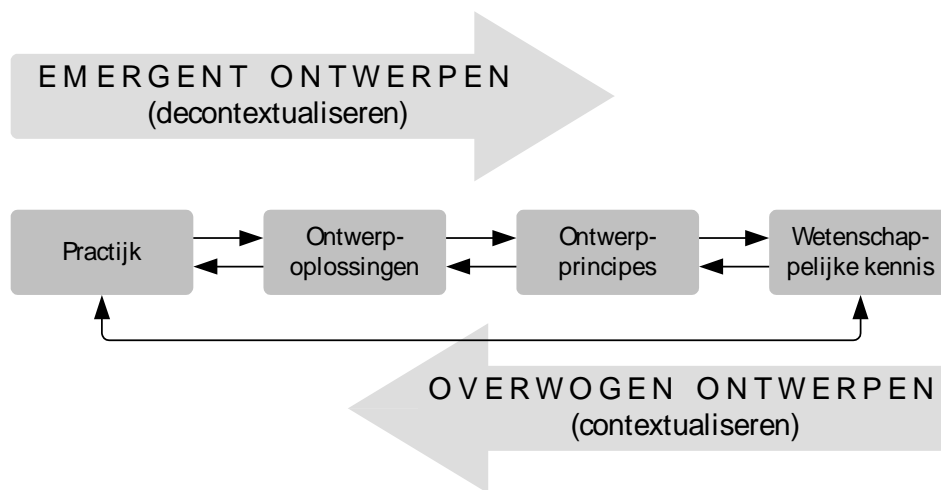
Illustratie: het ontwerpen van een infrastructuur voor universitaire spin-offs

Het gebruik van bovenstaande kwaliteitscriteria kan geïllustreerd worden met het onderzoek dat ik gedaan heb naar het ontwerpen van een infrastructuur voor spin-offs van universiteiten (Gilsing, Van Burg, & Romme, 2010; Van Burg, 2010; Van Burg et al., 2008). Spin-offs van universiteiten zijn bedrijven die opgericht worden om intellectueel eigendom van deze universiteiten te exploiteren. Deze spin-offs zijn in het bijzonder nuttig om doorbraken in technisch onderzoek, met potentieel dat anders ongebruikt zou blijven, om te zetten naar producten en diensten met marktwaaarde. Daarom zijn beleidsmakers erg geïnteresseerd in universitaire spin-offs, als middel voor overdracht van technologie en als mechanisme om economische groei te stimuleren. Mijn onderzoek naar spin-offs is erop gericht om beleidsmakers en managers op universiteiten ontwerpprincipes te geven die ze kunnen implementeren in hun locale context. Het onderzoek naar de ontwerpprincipes bestaat uit een veldstudie naar de praktijk op te Technische Universiteit Eindhoven en een systematisch literatuuronderzoek.

Ontwerpen als interactie tussen theorie en praktijk

In dit onderzoek naar het stimuleren van spin-offs zie ik het ontwerpproces als een interactie tussen praktijk en theorie. De resultaten van een ontwerpgericht onderzoek kunnen niet zomaar eenzijdig in een praktijksituatie toegepast worden, zonder interactie met die praktijk. Aan de andere kant zijn de inzichten en oplossingen die mensen in de praktijk – zonder al te veel theorie – hebben ontworpen nuttig voor het ontwerponderzoek en voor de algemene theorie.

In de alledaagse praktijk worden allerlei ontwerp oplossingen bedacht en uitgevoerd, vaak zonder daar veel over na te denken. Dat kun je 'emergent ontwerpen' noemen. Schön (1987) noemt dit 'reflection-in-action': midden in lopende proces bedenk je ineens een verbetering en je kijkt of die werkt. Als het werkt, zul je het 'herontwerp' de volgende keer en misschien ook in andere situaties toepassen. Aan de andere kant kun je vanuit bestaande (wetenschappelijke) kennis nieuwe ontwerp oplossingen aandragen en invoeren. Dit kun je 'weloverwogen ontwerpen' noemen. Figuur 1 laat beide vormen en hun interactie zien.



FIGUUR 1

Fig 9.1 De interactie tussen praktijk en wetenschap volgens een ontwerpgerichte benadering (gebaseerd op Van Burg et al. 2008)

Emergent ontwerpen en weloverwogen ontwerpen vormen de interactie tussen theorie en praktijk. Ontwerpmethodologie probeert deze twee werelden te overbruggen door ontwerpgeredenschap in het midden te plaatsen, met als belangrijk instrumentarium de ontwerpprincipes. Vanuit de wetenschappelijke kant moeten de inzichten dan meer gecontextualiseerd worden, om de algemene inzichten toe te passen op een specifieke organisatie, in een bepaalde sector, cultuur en regio (cf. Rousseau & Fried, 2001). Aan de andere kant moeten de praktische inzichten van hun specifieke context ontdaan moeten worden om de kennis ook voor andere situaties toepasbaar te maken. De kwaliteit van het ontwerpinstrumentarium en de daarbij horende contextualisatie en decontextualisatie wordt gewaarborgd door specifieke toepassingen van de kwaliteitscriteria voor onderzoek.

Gevalstudie

De gevalstudie had als doel om praktijkkennis te vertalen in een aantal ontwerpprincipes. De gegevens voor de veldstudie zijn in de periode 2005-2007 verzameld door de onderzoekers in twee verschillende rollen. Eén van de betrokken onderzoekers participeerde in het ontwerpproces binnen het TU/e Innovation Lab, de organisatie die binnen de universiteit verantwoordelijk was voor spin-off ondersteuning en technologietransfer. In de tweede rol, puur als onderzoeker, zijn er semi-gestructureerde interviews gedaan en zijn allerlei soorten documenten verzameld. Verschillende

betrokkenen vanuit de Technische Universiteit Eindhoven werden geïnterviewd: (potentiële) ondernemers van de spin-offs, managers van TU/e Innovation Lab, spin-off adviseurs, een adviseur voor intellectueel eigendom en verscheidene hoogleraren en docenten voor ondernemerschap. In totaal werden 25 interviews gedaan. Daarnaast heb ik zelf veel bijeenkomsten meegemaakt.

Deze combinatie van verschillende soorten respondenten en verschillende gegevensbronnen maakt het mogelijk om de stabiliteit van de onderzoeksresulten te beoordelen (betrouwbaarheid) en validiteit vast te stellen. De betrokkenheid van één van de onderzoekers in de TU/e Innovation Lab organisatie en mijn eigen observaties zorgen ervoor dat de gegevens niet alleen uit verbale of geschreven bronnen afkomstig waren, maar dat er ook observaties van handelingen en activiteiten waren (bevordering van pragmatische validiteit). Voor de interviews werd in alle gevallen dezelfde algemene vragenlijst gebruikt (zie Van Burg, 2010: 237-240) om de betrouwbaarheid (stabiliteit) van het onderzoek te bevorderen. De interviews werden letterlijk uitgetypt en gecodeerd volgens de procedures van Straus en Corbin (1998).

Om vanuit de interviews en documenten de principes te destilleren die in Eindhoven gebruikt werden voor spin-off ondersteuning, heb ik al het schriftelijk materiaal gedetailleerd gecodeerd (Plsek et al., 2007). De eerste fase van het coderen betrof de ervaringen en activiteiten die de ondernemers, adviseurs en andere geïnterviewden vertelden. Vervolgens werden deze codes geclusterd door gemeenschappelijke thema's te identificeren. De categorieën waarvoor te weinig onafhankelijke bronnen waren (minimaal drie) of die te weinig substantieel waren (bijvoorbeeld alleen uitspraken met aanbevelingen of ideeën, maar geen echte ervaringen of activiteiten) werden verwijderd. Dit leidde tot acht categorieën waarvoor voldoende materiaal was vanuit verschillende bronnen (stabiliteit en convergente validiteit) en dat onderling voldoende onderscheidend was (discriminante validiteit). Als laatste stap heb ik voor ieder van de overgebleven acht categorieën een samenvattend praktijk-gebaseerd ontwerpprincipe geformuleerd en bijbehorende voorbeelden beschreven. Twee voorbeelden van deze acht praktijk-gebaseerde principes zijn:

1. Creëer de benodigde fysieke infrastructuur voor spin-offs (zoals kantoorruimte) en sta toe dat spin-offs gebruik maken van de beschikbare infrastructuur van de universiteit (zoals laboratoria).
2. Zorg dat het ondernemerschapsonderwijs ingebed is in het curriculum en gericht is op het stimuleren van het oprichten van spin-offs.

Deze resultaten zijn daarna aan de geïnterviewden gepresenteerd om zo een 'member check' te doen (betrouwbaarheid).

Literatuursynthese

De bestaande wetenschappelijke kennis heb ik met een systematische literatuur review samengevat in een aantal ontwerpprincipes. Voor de opzet van deze review heb ik de methodologie van Denyer, Tranfield en collega's gevolgd (Denyer et al., 2008; Tranfield et al., 2003) en een eigen variant van het protocol ontwikkeld (betrouwbaarheid). Het domein voor deze review was: het onderzoek naar spin-offs van universiteiten. Het doel was om normatieve principes af te leiden, dus de review was niet in de eerste plaats bedoeld om een volledig overzicht van de literatuur te geven. Daarom heb ik in het reviewprotocol als startpunt een aantal samenvattende reviews van dit literatuurdomein genomen, namelijk Shane (2004), Djokovic en Souitaris (2008), en O'Shea, Allen, O'Gorman en Roche (2004). Omdat deze reviews in kwalitatief hoogstaande tijdschriften zijn verschenen en/of veel gerefereerd zijn, heb ik aangenomen dat de validiteit en betrouwbaarheid van deze reviews voldoende was.

Door de zogenaamde meta-analyse benadering te kiezen (zie hoofdstuk 11) kon ik de verschillende studies uit verschillende domeinen bij elkaar brengen (cf. Denyer & Tranfield, 2006; Tranfield, Denyer, Marcos, & Burr, 2004). Deze metha-synthese benadering is analoog aan de benadering van open coderen (Strauss & Corbin, 1998). De belangrijkste bron voor het coderen waren deze drie samenvattende reviews. Belangrijke bevindingen werden allemaal gecodeerd en indien nodig gerelateerd aan vergelijkbare concepten uit andere studies. Na het clusteren van de bevindingen resulteerde dit in twaalf categorieën, die door te zoeken naar overkoepelende thema's teruggebracht werden tot zes clusters. Om de convergente en discriminerende validiteit van deze zes categorieën te bevestigen, heb ik nog vijftien meer recente artikelen en working papers verzameld die nog niet in de drie kernreviews verwerkt waren. Hieruit bleek dat deze zes categorieën voldoende waren. Omdat de bevindingen uit de reviews echter nog niet altijd een goed verklarend principe opleverden (interne validiteit) en er vooral voor het 'generative mechanism' weinig bewijs was (construct validiteit), heb ik gezocht naar theorieën die de bevindingen uit de onderzoeken konden verklaren. Dat leverde een aantal 'generative mechanisms' op die verklarende kracht hadden. Uiteindelijk heb ik de resultaten geformuleerd in zes samenvattende literatuur-gebaseerde ontwerpprincipes. Twee voorbeelden van deze literatuur-gebaseerde ontwerpprincipes zijn:

1. Maak een selectie uit de technologieën en ideeën voor spin-offs en geef vervolgens deze spin-offs advies en coaching door ervaren mensen.
2. Maak duidelijke regels en procedures voor de exploitatie van kennis die van de universiteit is.

Ontwikkelen van uiteindelijke ontwerpprincipes

Vanuit de twee sets van praktijk- en literatuurgebaseerde ontwerpprincipes heb ik een uiteindelijke lijst met ontwerpprincipes ontwikkeld. Door de twee sets met ontwerpprincipes systematisch met elkaar te vergelijken en te zoeken naar overeenkomst en verschillen, ontdekte ik dat het mogelijk was om met vijf overkoepelende principes alle bevindingen samen te vatten (voor details zie Van Burg, 2010: 55-63; Van Burg et al., 2008). Twee voorbeelden van deze uiteindelijke ontwerpprincipes zijn:

Om de ontwikkeling van spin-offs vanuit de universiteit te bevorderen:

1. creëer bewustheid van de ondernemerschapsmogelijkheden op heel de universiteit, stimuleer het ontwikkelen van ondernemende ideeën en selecteer vervolgens de ondernemers en ideeën door programma's die gericht zijn op studenten, onderzoekers en docenten.
2. ondersteun teams van starters in het samenstellen van een team met de juiste mix van ondernemerschapsvaardigheden en -kennis door het bieden van toegang tot advies, coaching en training.

Voor verdere specificatie van de onderliggende mechanismes en voorbeelden van de praktische invoering van deze principes verwijst naar Van Burg (2010).

Deze uiteindelijke set met ontwerpprincipes is dus gebaseerd op zowel praktijk als theorie. Om de bruikbaarheid van deze uiteindelijke set van ontwerpprincipes te vergroten, heb ik voor ieder van de principes een lijst met voorbeelden gegeven van praktische uitwerking ervan. Daarnaast heb ik voor twee principes in detail beschreven hoe in Eindhoven de kennis vanuit deze ontwerpprincipes gebruikt is om daar de bestaande situatie mee te herontwerpen. Om de externe validiteit vast te stellen heb ik een overzicht gemaakt met kenmerken de TU Eindhoven case en dat vergeleken met kenmerken van andere universiteiten. Daaruit blijkt dat de resultaten waarschijnlijk transfereerbaar zijn naar andere Europese universiteiten met een focus op technologie of

Beta-vakken. Belangrijk is wel om vast te stellen of de wetgeving voor intellectueel eigendom overeenkomt.

Betrouwbaarheid en validiteit in dit onderzoek

In de bovenstaande beschrijving van de opzet, uitvoering en resultaten van dit onderzoek heb ik aangegeven welke maatregelen genomen zijn om betrouwbaarheid en validiteit te bewerkstelligen. Ter afsluiting wil ik kort een algemene evaluatie van betrouwbaarheid en validiteit geven en aangeven waar dit mogelijk nog verbeterd had kunnen worden.

Wat *betrouwbaarheid* betreft is de stabiliteit van de onderzoeksresultaten gewaarborgd door verschillende bronnen te gebruiken, de onderzoeksprocedures goed te beschrijven, member checks te gebruiken en de kwaliteit van de bronnen voor de literatuursynthese te controleren. De consistentie van de onderzoeksresultaten heb ik echter niet vast kunnen stellen. Een belangrijke taak voor vervolgonderzoek is dan ook om te kijken of deze principes ook uit andere cases naar voren komen. De onderzoeken die de in dit onderzoek ontwikkelde principes hebben gebruikt, lijken de consistentie te bevestigen (cf. Barr, Baker, Markham, & Kingon, 2009).

De validiteit van het onderzoek is op veel verschillende manier gewaarborgd. Construct validiteit door in het coderen van de gegevens en het formuleren van de principes te zorgen dat er voor alle onderdelen van het principe voldoende bewijs is (voor de context, de interventie, het mechanisme en de uitkomst). Voldoende bewijs voor ieder van de principes vereist dat het principe uit meerder onafhankelijke bronnen onderbouwd kan worden, zoals bijvoorbeeld uit zowel het literatuuronderzoek als uit het praktijkonderzoek. Convergente en discriminante validiteit zijn bevorderd en vastgesteld door gebruik te maken van verschillende bronnen en alles systematisch te coderen. Met betrekking tot interne validiteit is gezocht naar herhaling van het geobserveerde patroon in meerdere gegevensbronnen en naar het vinden van een logische en consistente verklaring in ieder ontwerpprincipie. Om de interne validiteit nog meer te versterken hadden rivaliserende verklaringen en herhaling van de resultaten in andere cases nog meer aandacht kunnen krijgen. De externe validiteit van dit onderzoek is alleen theoretisch bepaald, door een analytische generaliseerbaarheid vast te stellen. Wel hebben sommige onderzoeken de resultaten al toegepast in andere contexten, wat een bepaalde validatie van de generaliseerbaarheid geeft. De pragmatische validiteit is in dit onderzoek bevorderd door gebruik te maken van participant-observaties, door de resultaten helder te presenteren en door daarbij 'verhaaltjes' te geven over de toepassing van de principes. Om de pragmatische validiteit nog meer te versterken zou

vervolgonderzoek zich kunnen richten op het testen van het ontwerp en het ontwikkelen van overtuigende visualisaties.

Conclusie

In dit hoofdstuk staan de kwaliteitscriteria voor ontwerpgericht wetenschappelijk onderzoek centraal. Vanuit de algemene kwaliteitscriteria van betrouwbaarheid en validiteit is een specifieke vertaling gemaakt naar ontwerpgericht wetenschappelijk onderzoek. Ook geeft dit hoofdstuk een overzicht van de verschillende praktische maatregelen die gebruik kunnen worden om de betrouwbaarheid en validiteit van het onderzoek te verhogen. Hoe meer van deze methoden worden gebruikt, hoe hoger de kwaliteit van het onderzoek wordt. Met als uiteindelijke doel om onderzoek op te leveren dat niet alleen zelf van hoge kwaliteit is, maar ook tot goede en effectieve interventies leidt. In die zin is de kwaliteit van het onderzoek een belangrijke voorwaarde voor goede en effectieve interventies.

Referenties

- Argyris, C. (2006). *Reasons and rationalizations: The limits to organizational knowledge*. Oxford, UK: Oxford University Press.
- Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: The power of modularity*. Cambridge, MA: The MIT Press.
- Barr, S. H., Baker, T., Markham, S. K., & Kingon, A. I. (2009). Bridging the valley of death: Lessons learned from 14 years of commercialization of technology education. *Academy of Management Learning and Education*, 8(3), 370-388.
- Dammers, E. (2000). *Leren van de toekomst: Over de rol van scenario's bij strategische beleidsvorming*. Leiden: Eburon.
- Denyer, D., & Tranfield, D. (2006). Using qualitative research synthesis to build an actionable knowledge base. *Management Decision*, 44(2), 213-227.
- Denyer, D., Tranfield, D., & Van Aken, J. E. (2008). Developing design propositions through research synthesis. *Organization Studies*, 29(3), 393-413.
- Dijk, J. van, De Goede, M., 't Hart, H., & Teunissen, J. (1991). *Onderzoeken & veranderen: Methoden van praktijkonderzoek*. Houten: Stenfert Kroese.
- Djokovic, D., & Souitaris, V. (2008). Spinouts from academic institutions: A literature review with suggestions for further research. *Journal of Technology Transfer*, 33(3), 225-247.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Gilsing, V. A., Van Burg, E., & Romme, A. G. L. (2010). Policy principles for the creation and success of corporate and academic spin-offs. *Technovation*, 30(1), 12-23.

- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (Vol. 2, p. 163–194). Thousand Oaks, CA: Sage Publications.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 602-611.
- Kvale, S., & Brinkmann, S. (2009). *InterViews: Learning the craft of qualitative research interviewing*. Thousand Oaks, CA: Sage Publications.
- Lee, T. W. (1999). *Using qualitative methods in organizational research*. Thousand Oaks, CA: Sage Publications.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications.
- Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2008). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13.
- O'Shea, R., Allen, T., O'Gorman, C., & Roche, F. (2004). Universities and technology transfer: A review of academic entrepreneurship literature. *Irish Journal of Management*, 25(2), 11-29.
- Plsek, P., Bibby, J., & Whitby, E. (2007). Practical methods for extracting explicit design rules grounded in the experience of organizational managers. *Journal of Applied Behavioral Science*, 43(1), 153-170.
- Romme, A. G. L. (2003). Making a difference: Organization as design. *Organization Science*, 14(5), 558-573.
- Romme, A. G. L., & Endenburg, G. (2006). Construction principles and design rules in the case of circular design. *Organization Science*, 17(2), 287-297.
- Rousseau, D. M., & Fried, Y. (2001). Location, location, location: Contextualizing organizational research. *Journal of Organizational Behavior*, 22(1), 1-13.
- Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. San Fransisco: Jossey-Bass Publishers.
- Shane, S. (2004). *Academic entrepreneurship: University spinoffs and wealth creation*. Cheltenham: Edward Elgar.
- Simon, H. A. (1996). *The sciences of the artificial*. Cambridge: MIT Press.
- Stam, C. D. (2007). *Knowledge productivity: designing and testing a method to diagnose knowledge productivity and plan for enhancement*. Enschede: Universiteit Twente.
- Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Boston: Irwin/McGraw-Hill.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (Vol. 2). Thousand Oaks, CA: Sage Publications.
- Suh, N. P. (2001). *Axiomatic design: Advances and applications*. Oxford, UK: Oxford university press.
- Swanborn, P. G. (1996). A common base for quality control criteria in quantitative and qualitative research. *Quality and Quantity*, 30(1), 19-35.

- Taylor, C. (1985). *Philosophical papers 2: Philosophy and the human sciences*. Cambridge, UK: Cambridge University Press.
- Tranfield, D., Denyer, D., Marcos, J., & Burr, M. (2004). Co-producing management knowledge. *Management Decision*, 42(3/4), 375-386.
- Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207-222.
- Van Aken, J. E., Berends, H., & van der Bij, H. (2007). *Problem-solving in organizations*. Cambridge University Press.
- Van Aken, J. E., & Romme, A. G. L. (2009). Reinventing the future: Adding design science to the repertoire of organization and management studies. *Organization Management Journal*, 6(1), 2-12.
- Van Burg, E. (2010). *Creating spin-off: Designing entrepreneurship conducive universities*. Eindhoven: Eindhoven University Press.
- Van Burg, E., Romme, A. G. L., Gilsing, V. A., & Reymen, I. M. M. J. (2008). Creating university spin-offs: A science-based design perspective. *Journal of Product Innovation Management*, 25(2), 114-128.
- Wester, F., & Peters, V. A. M. (2004). *Kwalitatieve analyse: Uitgangspunten en procedures*. Bussum: Coutinho.
- Worren, N. A., Moore, K., & Elliott, R. (2002). When theories become tools: Toward a framework for pragmatic validity. *Human Relations*, 55(10), 1227 -1250.
- Yin, R. K. (2003). *Case study research: Design and methods* (Vol. third). London: Sage.